Point-to-point responses:

This manuscript aimed to extend previous framework of temporal variance decomposition in snow-dependent basins by incorporating the effects of snowmelt and vegetation changes. The topic is interesting and the manuscript is well structured. However, I have serious concerns with the methods and results (especially the robustness of the estimates of water cycle components) in the manuscript.

RESPONSE: We appreciate your constructive comments, and will revise the manuscript accordingly. In particular, to validate the reliability of the estimated water cycle components, we compared our results with previous studies, and found that our results are acceptable. Furthermore, the uncertainties were discussed.

4.5 Uncertainties

Uncertainties from different sources may result in errors for this study. First, this study estimated ΔS and Q_m with the GLDAS Noah land surface model and the degree-day model, respectively. Although the GLDAS_ΔS has been widely used in hydrological studies, it ignores the change in deep groundwater (Nie et al., 2016; Syed et al., 2008; Zhang et al., 2016), which may lead to errors in ET estimation based on water balance equation. But previous studies showed that the groundwater change in our study area is relatively small, and can thus be ignored. For example, Du et al. (2016) used the abcd model to quantitatively determine monthly variations of water balance for the sub-basins of Heihe River (including basins 3-5 in our study) and found that the soil water storage change have obvious effects on the monthly water balance, whilst the impact of monthly groundwater storage change is negligible. Furthermore, it has been found that any change in climate conditions and underlying basin characteristics will affect the contributions of heat balance components and cause temporal variations of DDF (Kuusisto, 1980; Ohmura, 2001). But previous studies indicated that there is no significant seasonal change in DDF in west China (Zhang et al., 2006); as such, it is acceptable to estimate snowmelt runoff using fixed DDF values in this study. In comparison, the contribution of snow meltwater to runoff (Fs) was 12.9% in Basin 2 during 1971-2015 by using Spatial Processes in Hydrology

model(Li et al., 2019), while Fs was 25% in Basin 3 from 2001 to 2012 based on geomorphology-based ecohydrological model (Li et al., 2018), <10% in Basin 6 during 1961-2006 by using SRM model (Gao et al., 2011). Our results indicated that the Fs in Basin 2, 3 and 6 were 14.8%, 24.5% and 6.7%, respectively, which were close to those from different models. Finally, the uncertainties of ΔS and Q_m may lead to errors in ET estimation by water balance equation. To validate the reliability of our estimated ET, the comparison with ET_{map} from April to September during 2012-2014 was conducted (Figure S4). The results showed that our estimated ET fitted well with ET_{map} and basically fell around the 1:1 line, indicating ET estimated using water balance equation by considering the items of ΔS and Q_m is acceptable.

Second, previous studies concluded that three main factors could be responsible for the variability of n, including underlying physical conditions (such as soil and topography characteristics) (Milly, 1994; Yang et al., 2007), climate seasonality (such as the temporal variability of rainfall, mismatch between water and energy) (Ning et al., 2017; Potter et al., 2005) and vegetation dynamics (Donohue et al., 2007; Zhang et al., 2001). On the short time scale, the changes in soil and topography are negligible and its impact on the variability of n can be ignored. In consequence, the factors, should be considered, are climate seasonality and vegetation dynamics. When parameterizing n, this study considered M but ignored climate seasonality since the covariance item between R and E_0, i.e. $\varepsilon_1\varepsilon_4 cov(R, E_0)$ in the attribution equation (13) can represent climate seasonality. In addition, human influence represented by parameter n on the water balance cannot be ignored, which remains further investigation.

The specific introduction of ET_{map} will be added in section 2.2-Data:

ET from dataset of "ground truth of land surface evapotranspiration at regional scale in the Heihe River Basin (2012-2016) ET_{map} Version 1.0" (hereafter "ET_{map}"), was used to validate the reliability of our estimated ET. This dataset was published by National Tibetan Plateau Data Center. It was upscaled from 36 eddy covariance flux tower sites (65 site years) to the regional scale with five machine learning algorithms, and then applied to estimate ET for each grid cell (1 km × 1 km) across the Heihe

River Basin each day over the period 2012–2016. It has been evaluated to have high accuracy (Xu et al., 2018). Basins 3,4,5 in our study belongs to the headwater sub-basins of Heihe River, and our monthly ET from April to September during 2012-2014 in these three basins was thus compared with ET_{map}.
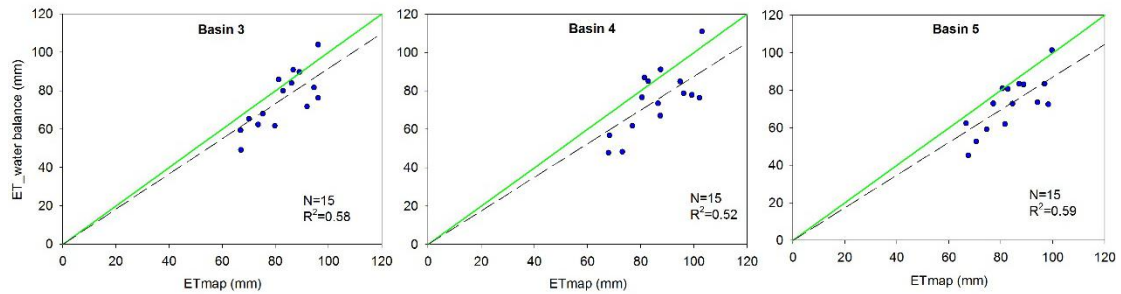

Figure S4. *Comparison of monthly ET derived from water balance equation and ET_{map} during 2012-2014.*

Comments

1. In this study, the total water storage is estimated using the GLDAS soil moisture and plant canopy surface water. Is this estimation reliable? More details about the methods (or additional comparison) may be needed to show the robustness of the total water storage estimation.

RESPONSE: The other reviewer also doubted the reliability of GLDAS-ΔS because it ignored the change in the deeper groundwater. For groundwater change, the best option is GRACE-ΔS; however, it is not applicable in the study area since the low spatial resolution of GRACE (1°×1°) would lead to large errors in small basins. Instead, GLDAS-ΔS is appropriate in representing the basin-scale water storage change. First, GLDAS has high spatial resolution of 0.25°×0.25°. Second, the groundwater change in west China is small and can be ignored. Specifically, Du et al. (2016) used the abcd model to quantitatively determine monthly variations of water balance for the sub-basins of Heihe River (including basins 3-5 in our study) and found that soil water storage change have effects on monthly water balance, whilst the impact of monthly groundwater storage change is negligible. To clarify the uncertainties, a new section will be added in the revised manuscript. The details can be found at the beginning.
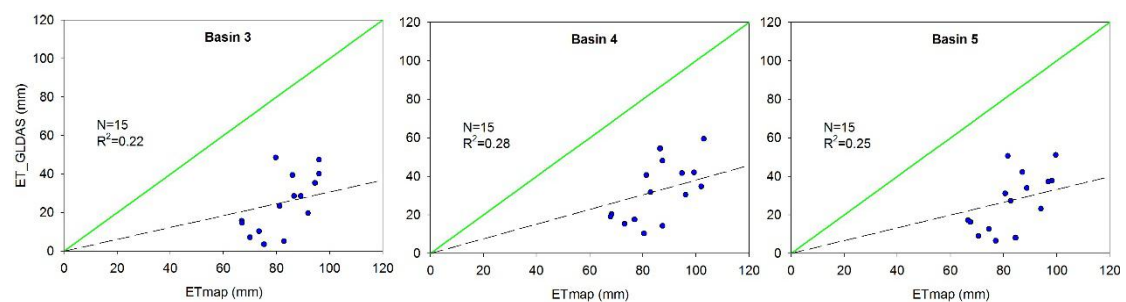
2. The degree-day model is used to estimate the equivalent of snowmelt runoff. In this model, the degree-day factors (DDF) in the study basins are fixed (if my understanding is correct here) and vary from 1.7-4.0 mm/day˚uC. Is there any uncertainty/validation of these factors? How the variation of the DDF could possibly affect the results of snowmelt runoff?

RESPONSE: Yes, your understanding is correct. It has been found that any changes in climate conditions and the underlying basin characteristics will affect the relative contributions of the heat balance components and cause temporal variations of the DDF (Kuusisto, 1980; Ohmura, 2001). But previous study indicated that there is no significant seasonal change in DDF in Western China (Zhang et al., 2006), which contains our study area. Thus, using the fixed DDF values to estimate snowmelt runoff is acceptable in this area. We also compared our snowmelt runoff values with other studies. In comparison, the contribution of snow meltwater to runoff (Fs) was 12.9% in Basin 2 during 1971-2015 by using Spatial Processes in Hydrology model(Li et al., 2019), while Fs was 25% in Basin 3 from 2001 to 2012 based on geomorphology-based ecohydrological model (Li et al., 2018), <10% in Basin 6 during 1961-2006 by using SRM model (Gao et al., 2011). Our results indicated that the Fs in Basin 2, 3 and 6 were 14.8%, 24.5% and 6.7%, respectively, which were close to those from different models. It can be concluded that our snowmelt runoff value is acceptable. Further, the uncertainties induced by the variation of the DDF will also discussed in revised manuscript. The details can be found at the beginning.

3. The total water storage and snowmelt runoff estimates are then used to calculate ET. Is the obtained ET reliable in terms of the above two comments?

RESPONSE: Even though there are many global ET products, but they have large uncertainty of its forcing data and model algorithms. Taking GLDAS-ET as an example, the precipitation data used come from the Princeton Global Fording dataset, which is a reanalysis dataset generated from a climate model. The spatial resolution is only 2°×2° (Sheffield et al., 2006). The low spatial resolution of forcing data should surely affect ET accuracy, especially in small basins. On the other hand, GLDAS ET products used Penman-Monteith equation to estimate ET. In this equation, the soil

water stress factor is critically important for plant transpiration suppression. However, this factor was implicitly considered by GLDAS products with the vapor pressure deficit (VPD). It is potentially problematic to use VPD to reflect soil water stress for transpiration, especially in drier regions. Some other promising recently released high-resolution ET products, such as GLEAM v3.2 and CLSM v2.0 also have similar problems. Thus, To validate the reliability of our ET, we conducted a comparison between our estimated ET, ET_GLDAS and ET from a dataset of "ground truth of land surface evapotranspiration at regional scale in the Heihe River Basin (2012-2016) ETmap Version 1.0", respectively. This ET dataset was published by National Tibetan Plateau Data Center. It was upscaled from 36 eddy covariance flux tower sites (65 site years) to the regional scale with five machine learning algorithms, and then applied to estimate ET for each grid cell (1 km × 1 km) across the Heihe River Basin each day over the period 2012–2016. It has been evaluated to have high accuracy. Basins 3,4,5 in our study belongs to the headwater sub-basins of Heihe River, and our monthly ET from April to September during 2012-2014 in these three basins was thus compared with ET_{map} (see Figure S4). The results showed that our estimated ET fits better with ET_{map} compared to GLDAS-ET and basically fell around the 1:1 line. Moreover, ET_GLDAS values is obviously smaller than ETmap. Even in the July and August, monthly ET_GLDAS is less than 60 mm, which is unreasonable in this region. Thus, it can be concluded that our estimated ET by water balance equation is acceptable. The details can be found at the beginning.

Comparison of monthly ET derived from GLDAS product and ETmap during 2012-2014.

4. I do not understand the results in Fig. 3. For example, we can see there are black dots in panel (b) (Pe=R-dS) with VERY low ET/Pe values (close to zero). If I

understand this correctly, when replace Pe with R+Qm-dS in panel (d), the ET/Pe should decrease as the Qm is positive (Table 1). It means that these low ET/Pe values in panel (b) should be more close to zero (close to x-axis) in panel (d). However, I did not see any black dots close to x-axis. WHERE are they? The results in Fig. 3 are confusing and do not make sense.

RESPONSE: Yes, as the Qm is positive, Pe with R+Qm-dS should increase compared with Pe with R-dS. But ET also increased, because it equals Pe-Qr according to equation (2) and (4). Thus, E0/Pe in x-axis will decrease while ET/Pe in y-axis will increase when considered Qm in panel d. Furthermore, the larger Qm, the larger increase in ET/Pe while decrease in E0/Pe, which means the black dots will move to the upper left after considering Qm.

The other reviewer also doubted the motivation of this part. Perhaps the simplest revised method is deleting the related content. But we thought it should be reserved as the following reasons: the Budyko framework was originally derived on long-term scale. Then it was gradually extended to characterize and predict the interannual variability of ET and the runoff fluxes on short time scale (including interanual and monthly scales). Some studies also showed that the Budyko framework was not suitable for exhibit ET variation on short time scale, because of the data points drew by ET ratio and dryness index beyond the two limit curves of Budyko framework (Chen et al., 2013; Du et al., 2016; Wang, 2012). These studies found that ignoring ΔS is the main reason (see Figure 11 by (Du et al., 2016); Figure 3 by (Chen et al., 2013)). Thus, validating the feasibility of using Budyko equation for variability of ET on the short time scale is the foundation.

Considering different combinations of water supply to ET is the main method for validation. In this study, except for ΔS, snowmelt runoff (Qm) is an important item of monthly water balance equation. Four combinations of water supply were thus assumed to prove the importance of considering ΔS and Qm into Budyko framework on monthly scale in the original manuscript. In this version, to avoid confusion, we only considered three combinations of water supply, i.e., Pe=R, Pe= R-ΔS and P_e=R-ΔS+Q_m.

Further, the related expression will also be revised.

4.1 The effects of monthly storage change and snowmelt runoff in the Budyko framework

The Budyko framework is usually used for analyses of long-term average catchment water balance; however, it was employed for the interpretation of the monthly variability of the water balance in this study. Thus, it's very necessary to validate the feasibility of Budyko equation for monthly variability. Furthermore, the impact of ΔS on the representation of Budyko framework on finer time scale has assessed by several studies (Chen et al., 2013; Du et al., 2016; Liu et al., 2019; Zeng and Cai, 2015). However, the impact of Q_m and its combined effects with ΔS in snowmelt-dependent basins are mostly ignored. Therefore, we present the water balance in the monthly scale of six basins in the Budyko's framework with three different computations of aridity index ($\phi=E_0/P_e$) or ET ratio (ET/Pe) in Figure 3. In Figure 3a, $ET=R-Qr$ when R is considered as water supply, i.e., $Pe=R$. The points of monthly ET ratio and aridity index in April and May were well below Budyko curves in 6 basins; monthly ET ratio was even negative in several year, which means the local rain are not the only sources of ET in this area, especially in spring. In Figure 3b, $ET=R-\Delta S-Qr$ with $Pe= R-\Delta S$. Compared with figure 3a, the way-off points in April and May were improved to a certain extent but negative points still existed, suggesting that except for R, ΔS also play a significant role in maintaining spring ET, but the variability of ET cannot be completely explained by these two variables. In Figure 3c, $ET=R-\Delta S+Qm-Qr$ with $P_e=R-\Delta S+Q_m$. Compared to the points in Figures 3a-b, all points focused on Budyko's curves more closely in each basin when $P_e=R+Q_m-\Delta S$ (Figure 3c). From this comparison, it can be concluded that the Budyko framework is applicable to the monthly scale in snowmelt-dependent basins, if the water supply is described accurately by considering ΔS and Q_m.
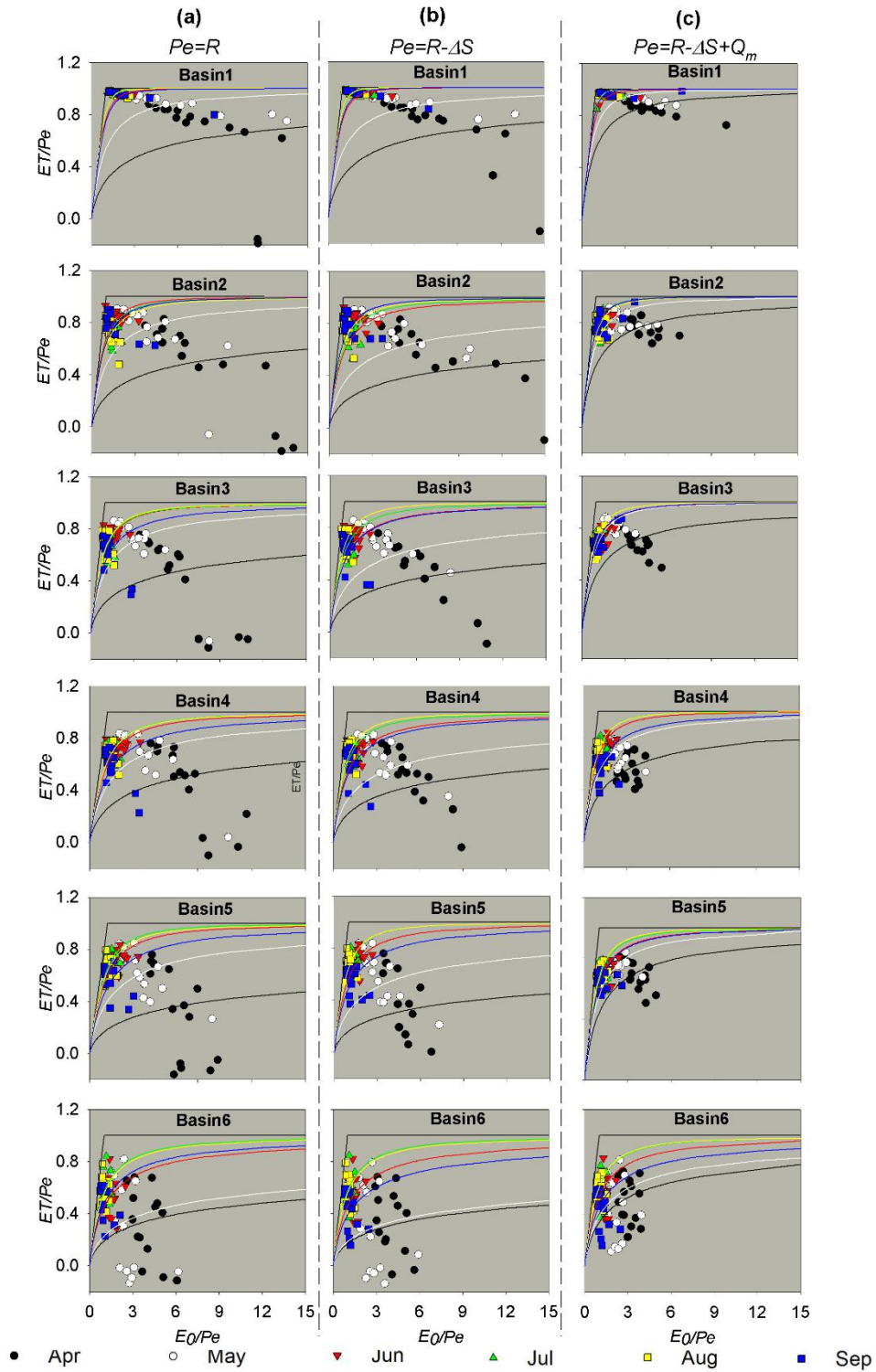
Figure 3 Plots for aridity index vs. evapotranspiration index scaled by available water supply for monthly series in growing season. Total water availability is (a) R, (b) R-ΔS and (c) R-ΔS +Qm. The n value for each Budyko curve is fitted by long-term averaged monthly data.

5. Do the Qs in the equations and the Qm in the figures have the same physical

meaning? If so, please keep the symbols consistent in the manuscript.

RESPONSE: We are so sorry for our carelessness. These two symbols all represented the snowmelt runoff. In the revised version, Qs will be revised as Qm.

6. In this manuscript, the term "temporal variance" is used in growing season by simply extending previous studies (e.g., Liu et al 2019). Is the definition of "temporal variance" in the growing season in this study the same as that in previous work? I cannot understand how it works in math.

RESPONSE: The definition of "temporal variance" in the growing season in this study is same as that in previous work. The only difference is the calculation of sample size (N) in equation 12. In previous studies, they focused on ET or runoff variance for all months. Thus, the sample size was 12 months/year × n years. In this study, we concerned ET variance in the growing season (April to September). Thus the sample size was 6 months/year × n years. The unbiased sample variance in equation 12 is estimated by the concept of statistics, not derived by previous studies or us. I would like to clarify the specific calculation as follows: in this study, with data of growing season (April to September) during 2001-2014, the sample size was 6 months/year×14 years=84 months, i.e. N=84 in equation 12. The calculation regarded all the months as a group or a time series of data, and did not conduct calculation for each calendar month. In consequence, i is used to index time series of month from 1 to N. $\overline{ET}$ is the long-term average of ET for 84 months. As such, one time series of data can only had one variance. It is known that a small test set size leads to a large bias in the estimate of the true variance between design sets (Geng et al., 1979; Wickenberg-Bolin et al., 2006). Comparing with conducting calculation for each calendar month, the calculation by us and other researchers (Liu et al., 2019; Ye et al., 2015; Zeng and Cai, 2015; Zeng and Cai, 2016) can obtain larger sample size. In the revised version, we will explain the related variables more clearly:

The unbiased sample variance of ET (σ_{ET}^2) is defined as:

$$\sigma_{ET}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(ET_i - \overline{ET})^2 = \frac{1}{N-1}\sum_{i=1}^{N}(\Delta ET_i)^2$$

(12)

where $\overline{ET}$ is the long term monthly mean of ET. N is the sample size, it equals 84 in this study (6 months/year×14 years=84 months). i is used to index time series of month from 1 to N.

References:

Chen, X., Alimohammadi, N.,&Wang, D., 2013. Modeling interannual variability of seasonal evaporation and storage change based on the extended Budyko framework. *Water Resources Research*, 49(9): 6067-6078.

Du, C., Sun, F., Yu, J., Liu, X.,&Chen, Y., 2016. New interpretation of the role of water balance in an extended Budyko hypothesis in arid regions. *Hydrology and Earth System Sciences*, 20(1): 393-409.

Gao, X., Zhang, S., Ye, B.,&Gao, H., 2011. Recent changes of glacier runoff in the Hexi Inland river basin. *Advances in Water Science (In Chinese)*, 22(3): 344-350.

Geng, S., Wang, W.J.,&Miller, C., 1979. SMALL SAMPLE-SIZE COMPARISONS OF TESTS FOR HOMOGENEITY OF VARIANCES BY MONTE-CARLO. *Communications in Statistics Part B-Simulation and Computation*, 8(4): 379-389.

Kuusisto, E., 1980. ON THE VALUES AND VARIABILITY OF DEGREE-DAY MELTING FACTOR IN FINLAND. *Nordic Hydrology*, 11(5): 235-242.

Li, H., Zhao, Q., Wu, J., Ding, Y., Qin, J., Wei, H.,&Zeng, D., 2019. Quantitative simulation of the runoff components and its variation characteristics in the upstream of the Shule River. *Journal of Glaciology and Geocryology (in Chinese)*, 41(4): 907-917.

Li, X., Cheng, G., Ge, Y., Li, H., Han, F., Hu, X., Tian, W., Tian, Y., Pan, X., Nian, Y., Zhang, Y., Ran, Y., Zheng, Y., Gao, B., Yang, D., Zheng, C., Wang, X., Liu, S.,&Cai, X., 2018. Hydrological Cycle in the Heihe River Basin and Its Implication for Water Resource Management in Endorheic Basins. *Journal of Geophysical Research-Atmospheres*, 123(2): 890-914.

Liu, J., Zhang, Q., Feng, S., Gu, X., Singh, V.P.,&Sun, P., 2019. Global Attribution of Runoff Variance Across Multiple Timescales. *Journal of Geophysical Research-Atmospheres*, 124(24): 13962-13974.

Nie, N., Zhang, W.C., Zhang, Z.J., Guo, H.D.,&Ishwaran, N., 2016. Reconstructed Terrestrial Water Storage Change (Delta TWS) from 1948 to 2012 over the Amazon Basin with the Latest GRACE and GLDAS Products. *Water Resources Management*, 30(1): 279-294.

Ohmura, A., 2001. Physical basis for the temperature-based melt-index method. *Journal of Applied Meteorology*, 40(4): 753-761.

Sheffield, J., Goteti, G.,&Wood, E.F., 2006. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, 19(13): 3088-3111.

Syed, T.H., Famiglietti, J.S., Rodell, M., Chen, J.,&Wilson, C.R., 2008. Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resources Research*, 44(2).

Wang, D., 2012. Evaluating interannual water storage changes at watersheds in Illinois based on long-term soil moisture and groundwater level data. *Water Resources Research*, 48: W03502.

Wickenberg-Bolin, U., Goransson, H., Fryknas, M., Gustafsson, M.G.,&Isaksson, A., 2006. Improved

variance estimation of classification performance via reduction of bias caused by small sample size. *Bmc Bioinformatics*, 7.

Xu, T., Guo, Z., Liu, S., He, X., Meng, Y., Xu, Z., Xia, Y., Xiao, J., Zhang, Y.,&Ma, Y., 2018. Evaluating Different Machine Learning Methods for Upscaling Evapotranspiration from Flux Towers to the Regional Scale. *Journal of Geophysical Research: Atmospheres*, 123: 8674-8690.

Ye, S., Li, H.-Y., Li, S., Leung, L.R., Demissie, Y., Ran, Q.,&Bloeschl, G., 2015. Vegetation regulation on streamflow intra-annual variability through adaption to climate variations. *Geophysical Research Letters*, 42(23): 10307-10315.

Zeng, R.,&Cai, X., 2015. Assessing the temporal variance of evapotranspiration considering climate and catchment storage factors. *Advances in Water Resources*, 79: 51-60.

Zeng, R.,&Cai, X., 2016. Climatic and terrestrial storage control on evapotranspiration temporal variability: Analysis of river basins around the world. *Geophysical Research Letters*, 43(1): 185-195.

Zhang, D., Liu, X., Zhang, Q., Liang, K.,&Liu, C., 2016. Investigation of factors affecting intra-annual variability of evapotranspiration and streamflow under different climate conditions. *Journal of Hydrology*, 543: 759-769.

Zhang, J., Liu, S.,&Ding, Y., 2006. Spatial Variation of Degree-day Factors on the Observed Glaciers inWestern China. *Acta Geographica Sinica (in Chinese)*, 61(01): 89-98.